Larimore

ESTIMATING THE NUMBER OF MARKED ANIMALS WHICH HAVE RETAINED THEIR IDENTITY FROM MULTIPLE MARKED ANIMALS AND ITS APPLICATION TO THE PETERSEN METHOD

by

Victor W. Lambou
Louisiana Wild Life and Fisheries Commission
Baton Rouge, Louisiana

and

Robert J. Muncy
School of Forestry
Louisiana State University
Baton Rouge, Louisiana

Presented at the Fifteenth Annual
Conference of the Southeastern Association
of Game and Fish Commissioners

October 22-25, 1961
Atlanta, Georgia

ESTIMATING THE NUMBER OF MARKED ANIMALS WHICH HAVE RETAINED THEIR IDENTITY FROM MULTIPLE MARKED ANIMALS AND ITS APPLICATION TO THE PETERSEN METHOD

Victor W. Lambou
Louisiana Wild Life and Fisheries Commission
Baton Rouge, Louisiana

and
Robert J. Muncy
School of Forestry
Louisiana State University
Baton Rouge, Louisiana

Abstract

In order to make an estimate of the size of a population of animals at a given time by the Petersen method, use is made of a sample of the fraction of marked animals in the population. However, if some of the animals originally marked lose their marks and thus can not be identified in the sample, a Petersen type estimate will be biased, the magnitude of the bias depending upon the proportion of animals retaining their identity. If an estimate can be made of the animals which have retained their identity at a given time, it is possible to make corrections for this bias. This report presents formulas for estimating the number of marked animals which have retained their identity at a given time from multiple marked animals and shows their derivation, shows their application to the Petersen method, discusses the necessary conditions for them to apply, discusses the errors associated with such estimates and shows how confidence limits can be determined.

Introduction

In order to make an estimate of the size of a population of animals at a given time by the Petersen method (or what game biologists commonly refer to as the Lincon Index) use is made of a sample of the fraction of marked animals in the population. The assumption is made that this sample fraction is an unbiased estimate of the fraction of marked animals in the population. However, if some of the animals originally marked lose their mark and thus can not be identified in the sample, a Petersen type estimate will be biased,

ESTIMATING THE NUMBER OF MARKED ANIMALS UNDER NAVE SAMPLING
UNDER DEPENDENT PROC RELATIVE MARKED ANIMALS AND ITS
APPLICATION TO WHITE PELICANS

Victor H. Lamkin
Louisiana Wild Life and Fisheries Commission
Baton Rouge, Louisiana

and

Robert L. Dorsey
Island of Presents
Louisiana State University
Baton Rouge, Louisiana

Abstract

In order to make an estimate of the size of a population of animals at a given time by the Petersen method, use is made of a sample of the recovery of marked animals in the population. However, if none of the animals originally marked lose their marks and can not be identified in the sample, a Petersen type estimate will be biased, the magnitude of the bias depending upon the proportion of animals retaining their identity. If an estimate can be made of the animals which have retained their identity at a given time, it is possible to make corrections for this bias. This report presents formulas for estimating the number of marked animals which have retained their identity at a given time from multiple marked samples and shows their consistency. Since their application to the Petersen method, discusses the necessary conditions for them to apply, discusses the errors associated with such estimates and shows how sensitive limits can be determined.

Introduction

In order to make an estimate of the size of a population of animals at a given time by the Petersen method for mark and recapture ecologists merely note, at the time marked are in made of a sample of the proportion of marked animals in the population. The assumption made that this sample therefore is an unbiased estimate of the proportion of marked animals in the population. However, if none of the animals originally marked lose their marks and this can not be identified in the sample, a Petersen type estimate will be biased.

the magnitude of the bias depending upon the proportion of animals retaining their identity. If an estimate can be made of the number of animals which have retained their identity at a given time, it is possible to make corrections for this bias. The purpose of this report is to present formulas for estimating the number of marked animals which have retained their identity at a given time from multiple marked animals and to show their derivation, to show their application to the Petersen method, to discuss the necessary conditions for them to apply, to discuss the errors associated with such estimates and to show how confidence limits can be determined. Also, it is suggested that more studies using techniques such as multiple marking be undertaken to determine how serious the bias from the loss of marks is in population estimates.

There appears to be considerable confusion relative to what is the correct method of estimating the number of marked animals which have retained their identity. For example, Murry (1960) used a procedure which gave a valid estimate of the number of marked deer that retained their identity; however, Dennett and Kidd (1960) used a procedure to estimate the number of marked animals which retained their identity which gave an estimate with a positive bias. It is hoped that this discussion in considering the essential points in population estimates in the case of multiple marking will illustrate what is the correct method.

Even though the previous studies were in reference to terrestial animals the same procedure applies to aquatic animals. Beverton and Holt (1957) in their book "On the Dynamics of Exploited Fish Populations" consider this problem and they especially cover the case where the loss of marks from multiple marked animals are not independent, which is not covered in this report.

The Petersen Method

A number of marked animals are placed in the population and then during some succeeding interval a sample of the population is taken to estimate the fraction of marked

the magnitude of the bias depending upon the proportion of animals retaining their identity. If an estimate can be made of the number of animals which have retained their identity at a given time, it is possible to make corrections for this bias. The purpose of this report is to present formulas for estimating the number of marked animals which have retained their identity at a given time from multiple marked animals and to show their derivation, to show their application to the Petersen method, to discuss the necessary conditions for them to apply, to discuss the errors associated with such estimates and to show how confidence limits can be determined. Also, it is suggested that more studies using techniques such as multiple marking be undertaken to determine how serious the bias from the loss of marks is in population estimates.

There appears to be considerable confusion relative to what is the correct method of estimating the number of marked animals which have retained their identity. For example, Darry (1963) used a procedure which gave a valid estimate of the number of marked mass that retained their identity; however, Bennett and Zidd (1960) used a procedure to estimate the number of marked animals which retained their identity which gave an estimate with a positive bias. It is hoped that this discussion in considering the essential points in population estimates in the case of multiple marking will illustrate what is the correct method.

Even though the previous studies were in reference to terrestial animals the same procedure applies to aquatic animals. Beverton and Holt (1957) in their book "On the Dynamics of Exploited Fish Populations" consider this problem and they especially cover the case where the loss of marks from multiple marked animals are not independent, which is not covered in this report.

The Petersen Method

A number of marked animals are placed in the population and then during some succeeding interval a sample of the population is taken to estimate the fraction of marked

animals in the population. The size of the population of animals can be estimated as follows:

$$t = M/(x/n) \qquad (1)$$

$$\text{or } t = Mn/x$$

where t = estimate of T, the size of the population at time of marking

M = number of animals originally marked

x = number of marked animals in the sample

n = number of animals in the sample

In the notation employed in this report a capital letter usually refers to a population parameter and a small letter refers to a sample parameter or estimate from the sample.

Formula (1) is the maximum - likelihood estimate of the population and it is consistent in the sense that it tends to the correct value as the sample fraction is increased. However, it is biased, and in small samples this bias can be substantial (Bailey, 1952). This bias is positive and of the order f^{-1} where $f = E(x) = Mn/T$ (Bailey, 1952). However, in large enough samples this bias will not be serious. Bailey proposed a modified formula which gives an almost unbiased estimate. Thus:

$$t = M(n+1)/x+1 \qquad (2)$$

Bailey (1952) and others have pointed out that the estimate of the reciprocal of the populations size $Y = T^{-1}$ is unbiased. Thus:

$$y = (x/n)/M \qquad (3)$$

$$\text{or } y = x/nM$$

where $y = t^{-1}$, the estimate of the reciprocal of the population size.

Also, according to Ricker (1958) the reciprocal has the advantage that it has the tendency to be distributed more nearly symmetrically about the population mean.

animals in the population. The size of the population of animals can be estimated as follows:

$$\hat{y} = M/(x/n) \qquad (1)$$

or $\hat{y} = Mn/x$

where $\hat{y}$ = estimate of y, the size of the population at time of marking

M = number of animals originally marked

x = number of marked animals in the sample

n = number of animals in the sample

In the notation employed in this report a capital letter usually refers to a population parameter and a small letter refers to a sample parameter or estimate from the sample.

Formula (1) is the maximum - likelihood estimate of the population, and it is equivalent in the sense that it tends to the correct value as the sample fraction is increased. However, it is biased, and in small samples this bias can be substantial (Bailey, 1952). This bias is positive and of the order y^{-1} where $y = E(\hat{y}) = Mn/x$ (Bailey, 1952). However, in large enough samples this bias will not be serious. Bailey proposed a modified formula which gives an almost unbiased estimate. Thus:

$$\hat{y} = M(n+1)/(x+1) \qquad (2)$$

Bailey (1952) and others have pointed out that the estimate of the reciprocal of the population size $Y = y^{-1}$ is unbiased. Thus:

$$\hat{Y} = (x/n)/M \qquad (3)$$

or $\hat{Y} = x/Mn$

where $Y = y^{-1}$, the estimate of the reciprocal of the population size.

Also, according to Madow (1939) the reciprocal has the advantage that it has the tendency to be distributed more nearly symmetrically about the population mean.

Ricker (1958) lists the necessary conditions for the application of formulas (1) through (3). These are:

(1) The marked animals suffer the same mortality as the urmarked.

(2) The marked animals are as vulnerable to recapture as the unmarked animals.

(3) The marked animals do not lose their mark.

(4) The marked animals become randomly mixed with the unmarked animals or the subsequent sample is proportional to the number of animals present in different parts of the population.

(5) All marked animals are recognized and reported on recovery.

(6) There is only a negligible amount of recruitment to the catchable population during the time the recoveries are being made.

The satisfaction of all of the above conditions is important in the straight forward application of the Petersen method. However, this paper is concerned mainly with condition (3) and how to correct for it when it is not satisfied.

Development of Sampling Model

In almost all of the development to follow, use will be made of the concept of compound probabilities of independent events. Therefore this concept will be introduced at the start.

If two independent events have probabilities of occurrence of $P(a_1)$ and $P(a_2)$ respectively, then the compound probability $P(a_1, a_2)$ that both events will occur together is the product of their separate probabilities. Thus,

$$P(a_1, a_2) = P(a_1)\, P(a_2)\,. \tag{4}$$

Now if $P(a_1) = P(a_2)$, then,

$$P(a_1, a_2) = P(a)^2, \tag{5}$$

If more than two independent events are involved then,

$$P(a_1, a_2, \ldots, a_r) = P(a_1)\, P(a_2) \ldots P(a_r) \tag{6}$$

Digitized by the Internet Archive
in 2023 with funding from
University of Illinois Urbana-Champaign Alternates

https://archive.org/details/estimatingnumber00lamb

and if $P(a_1) = P(a_2) = \ldots = P(a_r)$, then,

$$P(a_1, a_2, \ldots, a_r) = P(a)^r. \qquad (7)$$

This can be verified in almost any introductory text dealing with probability, such as Goldberg (1960) or in more advanced texts such as Feller (1957) and Anderson and Bancroft (1952).

We are assuming that an animal is marked with mark 1, mark 2 through r marks and that the event (a), the loss of a mark, occurs independently and randomly throughout the population and that the events (a_1), (a_2), . . ., (a_r), the loss of mark 1, 2, . . ., r, respectively, are mutually independent.

Assuming that our assumptions apply, then our problem is to determine the number of animals which lose all r marks, i.e., the number of animals which lose their identity. If we know the probabilities of losing mark 1, 2, . . ., r, then from formulas (4) through (7), we can determine the probabilities of losing all r marks. Probability is sometimes defined as the proportion in a large number of repeated independent trials. Then, the number of animals expected to lose all marks would be $MP(a_1, a_2, \ldots, a_r)$ where M is the number of animals originally marked. However, the probabilities of losing a mark is not known, but these can be estimated from the sample. Such estimates are known as emperical probabilities.

Actually, we are dealing with random variables and we are interested in the joint probability distribution of mark 1, 2, . . ., r. Let's consider the joint distribution of the loss or retention of mark 1 and mark 2. This joint distribution will be a bivariate frequency function represented symbolically by $f(m, \bar{m})$ where m refers to mark 1 and $\bar{m}$ refers to mark 2. Let event (a), the loss of a mark be represented by 0 and event b the retention of a mark be represented by 1, then their joint probability distribution would be as follows:

$\bar{m}$ \ m	m_o	m_1	Totals
$\bar{m}_o$	$f(\bar{m}_o, m_o)$	$f(\bar{m}_o, m_1)$	$f(\bar{m}_o)$
$\bar{m}_1$	$f(\bar{m}_1, m_o)$	$f(\bar{m}_1, m_1)$	$f(\bar{m}_1)$
Totals	$f(m_o)$	$f(m_1)$	1

From the above it can be seen that $\sum_{W} f(\bar{m}_i, m_i) = 1$, where W is the entire $(\bar{m}, m)$ region. However, we are interested in a subregion, w of W, that is, the region $(\bar{m}_o, m_o)$. Since $\sum_{w} f(\bar{m}_i, m_i)$ can be computed for any subregion, then $P(w) = \sum_{w} f(\bar{m}_i, m_i)$ which in this case would be the probability of losing both marks. Then the probability of not losing both marks, which is the subregion not included in w, is $1-P(w)$. Note, the totals of the columns give the probability distribution of $\bar{m}$ and the totals of the rows give the probability distribution of m, which are designated as the marginal distributions of $\bar{m}$ and m. Also note the $\sum f(\bar{m}_i)$ and the $\sum f(m_i)$ both equal 1 and that in this case they will be binomially distributed. The $\sum_{W} (\bar{m}_i, m_i)$ will also equal 1.

Since $\bar{m}$ and m are independent $f(\bar{m}_i, m_i) = f(\bar{m}_i)\, f(m_i)$ and the entry in any row and column is the product of $f(\bar{m}_i)$ and $f(m_i)$ and which will give the probability of event $\bar{m}_i$ and m_i occurring together. The function, $f(\bar{m}_o, m_o)$ is also a random variable and since we are interested only in whether an animal loses both marks or does not lose both marks, this function will be distributed binomially, i.e., $\left[P(w) + P(q)\right]^n = 1$, where $P(q) = 1 - P(w)$.

From the previous we can devise a sampling plan to estimate the number of animals retaining their identity. If we take a random sample, with replacement, of the animals

with mark 1, then, since we are sampling from the binomial probability distribution, an unbiased estimate of $P(a_1)$, the probability of losing mark 1 is

$$p_1 = \frac{a_1}{n} \tag{8}$$

where p_1 = the proportion of the animals in the sample losing mark 1

a_1 = the number of animals in the sample which lost mark 1

n = the number of animals in the sample

and the expected value of p_1 is

$$E(p_1) = E(a_1/n) \tag{9}$$

Likewise, $P(a_2)$, the probability of losing mark 2 is estimated by

$$p_2 = a_2/n \tag{10}$$

and

$$E(p_2) = E(a_2/n) \tag{11}$$

Then, $P(a_1, a_2)$ can be estimated by

$$p_{1,2} = p_1 p_2 \tag{12}$$

where $p_{1,2}$ = the estimate of the proportion of animals originally marked with mark 1 and 2, which lost both mark 1 and 2

and

$$E(p_{1,2}) = E(p_1)\, E(p_2) \tag{13}$$

Therefore, $p_{1,2}$ is an unbiased estimate of the proportion of animals losing their identity in an infinite population. However, this is only the most likely proportion of animals losing their identity in the finite population of animals we have marked (which can be considered as a sample from an infinite population). It is not necessarily the proportion losing their identity in the finite population, even if the proportion of animals losing

mark 1 and mark 2 was known without error. If $p_{1,2}$ is an unbiased estimate, then $1-p_{1,2}$ is an unbiased estimate of the proportion of animals retaining their identity.

So far we have considered only the joint distribution of the loss or retention of mark 1 and mark 2; however, this can be extended in a similar manner to 3 or more marks.

Estimating Animals Which Retained Their Identity

If a number of multiple marked animals are put into the population and then during some succeeding interval a sample of the population is taken then it is possible to estimate the number of marked animals which have retained their identity. Let the animals be originally marked with two marks, mark 1 and mark 2. Let's take an unbiased sample of the animals and assume that the loss of mark 1 is independent of mark 2 and that the loss of marks occurs randomly throughout the population. In the sample some animals will be found to have lost mark 1 and mark 2. All animals marked with mark 1 will be an unbiased sample of those which were originally marked with mark 2 some of which lost mark 2 and some of which retained mark 2. This would have to be true by definition, since the conditional probability of losing mark 1 given that mark 2 is not lost is equal to the probability of losing mark 1 and the conditional probability of losing mark 2 given that mark 1 is not lost is equal to the probability of losing mark 2, if the loss of mark 1 and 2 are independent of each other.

Then an estimate of the proportion of animals losing mark 2 is

$$p_2 = \frac{a_2}{b_1} \tag{14}$$

where p_2 = estimate of the proportion of animals losing mark 2

a_2 = the number of animals in the sample which lost mark 2

b_1 = the number of animals in the sample with mark 1

Likewise,

$$p_1 = \frac{a_1}{b_2} \tag{15}$$

where p_1 = estimate of proportion of animals losing mark 1

a_1 = the number of animals in the sample which lost mark 1

b_2 = the number of animals in the sample with mark 2.

Then $p_{i, \ldots, r}$) an estimate of the proportion of animals losing their identity or proportion of animals losing the ith through rth tag is,

$$p_{1,2} = p_1 p_2 \tag{16}$$

and an estimate of the proportion of animals not losing their identity, is

$$q = 1 - p_{i, \ldots, r} \tag{17}$$

Then s, an estimate of the number of animals originally marked which have retained their identity, is

$$s = qM \tag{18}$$

So far, we have been assuming that $P(a_1) \neq P(a_2)$; however, if we can safely assume that the probability of losing mark 1 is equal to the probability of losing mark 2, i.e., $P(a_1) = P(a_2)$, it will be better to pool the sampling data and make one estimate of P(a) the probability that any mark will be lost. Then, p, the estimate of the proportion of marks lost is

$$p = \frac{a_2 + a_1}{b_2 + b_1} \tag{19}$$

An equivalent of formula (19), which probably is the form that can be used the easiest with field data is

$$p = \frac{c}{2d - c} \tag{20}$$

where c = the number of animals in the sample which lost one mark

d = number of marked animals in the sample

Then

$$p_{1,2} = p^2 \tag{21}$$

and an estimate of the number of animals originally marked which have retained their identity can be made by using formulas (11) and (12).

Now let the animals be originally marked with three marks, mark 1, mark 2, and mark 3. Let's take an unbiased sample of the animals and assume that the loss of mark 1, 2 and 3 are independent of each other. Then an animal marked with mark 2 and/or mark 3 will be an unbiased sample of those which were originally marked with mark 1. An estimate of the proportion of animals losing mark 1 is

$$p_1 = \frac{a_1}{b_{2,\,3}} \tag{22}$$

where a_1 = the number of animals in the sample losing mark 1

$b_{2,\,3}$ = the number of animals in the sample with mark 2 and/or mark 3

Likewise,

$$p_2 = \frac{a_2}{b_{1,\,3}} \tag{23}$$

$$\text{and } p_3 = \frac{a_3}{b_{1,\,2}} \tag{24}$$

where a_2 = the number of animals in the sample losing mark 2

a_3 = the number of animals in the sample losing mark 3

$b_{1,\,3}$ = the number of animals in the sample with mark 1 and/or mark 3

$b_{1,\,2}$ = the number of animals in the sample with mark 1 and/or mark 2

Then,

$$p_{1,2,3} = p_1 p_2 p_3 \tag{25}$$

If we can safely assume that $P(a_1) = P(a_2) = P(a_3)$, then p the estimate of the proportion of marks lost is

$$p = \frac{a_1 + a_2 + a_3}{b_{2,3} + b_{1,3} + b_{1,2}} \tag{26}$$

and

$$p_{1,2,3} = p^3 \tag{27}$$

It should be noted that it is possible to assume that $P(a_1) = P(a_2) \neq P(a_3)$. Then we should pool the data for the estimate of the proportion losing mark 1 and 2 into one sample. Also it is possible to use more than three marks, ... which can be handled in a manner similar to the above examples.

Discussion of Problems Concerned with Estimating Number of Animals Which Have Retained Their Identity

In taking our subsequent sample of the marks and estimating the number of animals which have retained their identity there are possible sources of bias which could enter into our estimate even though our sample was random. If the sample is taken over a short period of time where the increase in the number of marks being lost is negligible, then any possible bias would be negligible. However, if the increase in the number of marks being lost is substantial, it is possible that the estimate of the number of animals which have lost their identity will be biased. During the sampling period we are estimating the average number of marks retained, and, we are making a linear interpolation. However, if during the sampling period the graph of the number of marks retained plotted against time departs any appreciable amount from linearity, then our estimate would be biased. However, if the graph is linear or essentially linear, there should be no appreciable bias.

As mentioned previously, if we can safely assume that $P(a_1) = P(a_2)$ or $P(a_1) = P(a_2) = \ldots = P(a_r)$ then obviously it will be better to pool the sampling data and make one estimate of $P(a)$. However, the question arises, when is it safe to assume that the probabilities are equal? One could test to see if the observed proportions of marks lost differ significantly by (1) a chi-square test of independence (2) a binomial test or (3) if the sample size is small more accurately from statistical tables for use with binomial samples such as presented by Mainland (1956). Then if the differences are significant, we would decide against pooling. However, in tests of this nature, one can make the statement that the differences are non-significant or are of such a magnitude as to be inconsequential, but it is impossible to prove that the differences are zero. It may be that our tests are not accurate enough to detect the differences. Also our tests are for Type I errors (Type I error is to reject the null hypothesis when it is true). However, in a test of this nature it would appear that a Type II error (Type

II error is to accept the null hypothesis when it is false) is much more serious than a Type I error. Therefore, in order to minimize the probability of committing a Type II error, we believe it would be wise to accept differences as being significant at a relatively high probability (say, e.g., probability of .30 which is the probability of committing a Type I error). This is possible, inasmuch as when the sample size is fixed, choosing as significant a larger probability of a Type I error will decrease the probability of making a Type II error. Generally, we would recommend that one assume that $P(a_1) \neq P(a_2)$ or $P(a_1) \neq P(a_2) \neq \ldots \neq P(a_r)$, unless there is strong evidence to the contrary. Estimates of Q will not be biased by assuming that the probabilities are different, but our estimate of Q could be biased by assuming that they are equal when in fact they are not equal.

It should be pointed out that in order to estimate Q it is not necessary to multiple mark all animals, i.e., if we can identify the multiple marked animals in the sample. Also we could estimate Q from a separate experiment, during a different time period, a different area, or even by pooling data from various sources. Obviously there is inherent danger in such a procedure. One should be certain that the estimate of Q applies to the period for which we are estimating the number of animals retaining their identity.

This paper has the objective of showing how to estimate the number of animals which have retained their identity so that a correction can be made for this in estimating population size. However, we would like to point out certain aspects which should be considered even if an estimate of the number of animals retaining their identity is not made. With some animals, e.g., deer, the cost of capturing the animals for marking is high. Therefore, once such an animal is marked and released into the population we want to be relatively certain that we will be able to identify this animal again. Therefore, there will be obvious advantages in multiple marking such an

animal. Usually the cost of marking and handling the animal is negligible compared to the cost of capture. Also, the cost of applying any additional marks on the animal is negligible. Previous experiments may indicate that a mark placed on such an animal will after a given period of time, say a year, have a probability of being lost equal to .10. Then we would expect that on the average 10 percent of our animals would lose their identity while 90 percent would retain their identity after a year's time if only one mark was placed on the animals originally. However, if two marks were placed on each animal and each mark had a probability of being lost of .10, then $p_{1,2} = .10^2 = .01$. Thus, we would expect that on the average only 1 percent of the animals would lose their identity while 99 percent would retain their identity. Now, let's place 3 marks on the animal each with a probability of being lost of .10; then $p_{1,2,3} = .10^3 = .001$. Thus we would expect that on the average only .1 percent of the animals would lose their identity while 99.9 percent would retain their identity. This is a considerable reduction from when only one mark was placed on the animal. The multiple marking would be another approach to overcoming the problems of mark loss and may be more fruitful than attempts toward obtaining superior marks.

With animals which are small and can be caught in large numbers it would not be as important to make sure that the animals have such a high probability of retaining their identity, i.e., if we can evaluate this loss and correct for it in our population estimate. In fact, the multiple marking of such animals might have a disadvantage. If marking and handling cause mortality, it is possible that the use of multiple marks will cause additional mortality. This could bias our population estimate, inasmuch as stated previously, one of the necessary conditions for the application of the Petersen estimate is that marked animals suffer the same mortality as the unmarked.

Estimating the Population

Once we have determined s, the estimate of the number of animals originally marked which have retained their identity, we are in a position to make a more nearly unbiased estimate of the population by the Petersen method. In the Petersen formulas, formulas (1), (2), and (3), M is the number of animals originally marked. However some of these animals will not be available to be caught as marked animals in the sample because they have lost their identity. Therefore if we substitute s for M in the Peterson formulas we will have a more nearly unbiased estimate of the population. Thus, formula (1) becomes

$$t = s/(x/n)$$

$$\text{or} \quad t = sn/x \tag{28}$$

and formula (2) becomes

$$t = x(n+1)/x+1 \tag{29}$$

and formula (3) becomes

$$y = (x/n)/s$$

$$\text{or} \quad y = x/ns \tag{30}$$

We have been assuming that the probability of the loss of marks on an animal are independent, i.e., their loss is not correlated. It will be worthwhile to consider the consequences if they are positively correlated, i.e., the loss of one mark increases the probability that another mark will be lost. (It would appear that the possibility of them being negatively correlated is unlikely.) Then our estimate, $p_{i,\ldots,r}$, will have a negative bias while our estimate of Q will have a positive bias, i.e., on the average $q > Q$. Then on the average $s > S$ and on the average our population estimate will have a positive bias, i.e., $t > T$. However, if we did not estimate s (even though s is biased) our population estimate would on the average have an even greater bias, i.e., on the average $t_1 > t_2 > T$, where t_1 is the population estimate when no estimate is made of the number of animals retaining their identity and t_2 is the population estimate where an

estimate is made of the number of animals retaining their identity. Thus, even though the losses of marks are positively correlated it will be worthwhile to make a biased estimate of Q, for it will always allow us to remove some even though not all of the bias from our estimate of T.

SAMPLING ERROR

Any estimate is subject to experimental error and it is important to make some statement about the probable size of such error. The size of such error must be evaluated, at least approximately, before any confidence can be placed in an estimate. In the usual Petersen type estimate there will be sampling error in x (or x/n) and in formulas (28), (29) and (30) there will also be sampling error in our estimate of s. If our assumptions are met, all of the quantities to be estimated will be distributed according to the binomial distribution, if sampling is done with replacement. However, if sampling is done without replacement they will be distributed according to the hypergeometric distribution. However, in practice, it would rarely happen that so great a fraction of the population is sampled that the hypergeometric would differ appreciably from the binomial. Thus, in most instances no sensible error would be committed by taking the distribution to be binomial. We are also considering that our sample size is large enough and that the mean of our estimate of the population size, t or y, is large enough for large sample theory to apply and so that t and y may be regarded as normal variables. Then the 95 percent confidence limits of our estimate (actually the 95.4 percent limits) can be approximated by

$$(\underline{t}, \overline{t}) = t \pm 2\sqrt{v(t)} \qquad (31)$$

where $\underline{t}$ = the lower limit of the 95 percent confidence interval

$\overline{t}$ = the upper limit of the 95 percent confidence interval

$v(t)$ = the sample estimate of the variance of t.

and $(\underline{y}, \overline{y})$ can be estimated in a similar manner.

In calculating confidence limits as outlined above, they should preferably be calculated for statistics whose distribution is as "normal" as possible.

Ricker (1958) states that in estimating the population size, the reciprocal of t tends to be distributed symmetrically about the mean, while t often is not symmetrically distributed about the mean. Therefore, he suggests that confidence limits first be computed for y (where $y = t^{-1}$) and then inverted in order to obtain limits for t.

According to the binomial theory, the large sample estimate of the variance of a proportion is:

$$v(p') = p'(1-p')/n \tag{32}$$

where $v(p')$ = the sample estimate of the variance of p'

p' = the sample estimate of the proportion

n = sample size

In many of the estimates discussed in this report there will be experimental error in two or more components and it will be necessary to compound these errors. Therefore we will present formulas for the large sample approximation of the variance of a product and the variance of a quotient. If $Z = XY$, then the variance of Z is

$$V(Z) = Y^2V(X) + X^2V(Y) \tag{33}$$

where $V(Z)$ = the variance of Z

$V(X)$ = the variance of X

$V(Y)$ = the variance of Y.

Now if $Z = Y/X$, then the variance of Z is

$$V(Z) \cong \frac{Y^2}{X^4} V(X) + \frac{1}{X^2} V(Y) . \tag{34}$$

In sampling, we substitute estimates from the sample into the above formulas and thus we can arrive at an estimate of the variance of the product or quotient. The development of these formulas are shown in appendix (1).

To illustrate the setting of confidence limits, let's illustrate the procedure for computing confidence limits for formula (1) and formula (3) assuming that none of the marked animals lose their identity and M is known without error. The variance of x/n from formula (32) is

$$v(x/n) = (x/n)(1-x/n)/n. \tag{35}$$

Then from formula (34) the variance of t is

$$v(t) = \frac{M^2}{(x/n)^4} v(x/n) + \frac{1}{(x/n)^2} v(M) .$$

However, since v(M) is zero the term $\frac{1}{(x/n)^2} v(M)$ vanishes and

$$v(t) = \frac{M^2}{(x/n)^4} v(x/n) . \tag{36}$$

The variance of y from formula (34) is

$$v(y) = \frac{(x/n)^2}{M^4} v(M) + \frac{1}{M^2} v(x/n) .$$

However, since v(M) is zero, the term

$$\frac{(x/n)^2}{M^4} v(M)$$

vanishes and

$$v(y) = \frac{1}{M^2} v(x/n) . \tag{37}$$

Bailey (1952) gives the following expression for estimating the variance of t:

$$v(t) = M^2 n(n-x)/x^3 . \tag{38}$$

Ricker (1958) gives the following expression for the large sample variance of y:

$$v(y) = \frac{x(n-x)}{M^2 n^3} . \tag{39}$$

However, these expressions (38) and (39) will give the same results as formulas (36) and (37). This can be illustrated by the following example taken

from Ricker (1958, p. 85, example 3A). Thus, M = 109, n = 177 and x = 57. By formulas (1) and (3), t = 109/(75/177) = 339 and y = (57/177)/109 = .00295. The variance of x/n is:

$$v(x/n) = (.322)(.678)/177 = .001233$$

and according to formula (36)

$$v(t) = \frac{109^2}{.322^4} \quad .001233$$

$$= 1362.677 \quad .$$

According to formula (38)

$$v(t) = (109)^2 \ (177)(177-57)/57^3$$

$$= 1362.646$$

which gives the same result as formula (36) allowing for errors due to rounding. According to formula (37)

$$v(y) = \frac{1}{109^2} \quad .001233$$

$$= .000,000,010,38$$

and according to formula (39)

$$v(y) = \frac{57(177-57)}{(109)^2(177)^3}$$

$$= .000,000,010,38$$

which is the same result as obtained from formula (37).

The approximate 95 percent confidence limits of t and y are

$$(\underline{t}, \overline{t}) = 339 \pm 2\sqrt{1362.677}$$

$$= 265\text{-}413$$

and $$(\underline{y}, \overline{y}) = .00295 \pm 2\sqrt{.000,000,010,38}$$

$$= .002,747 - .003,153$$

Inverting the limits of y gives limits of $\underline{t}$ and $\overline{t}$ equal to 317 - 364 which are not symmetrical about the mean.

Now let's consider the case where some of the marked animals lose their identity and we are using formulas (28) and (30) to estimate the population. The variance of (x/n) will be estimated as done previously (formula 35). However, now there is also experimental error in our estimate of s. First we must estimate the variance of p or $p_1 p_2$, etc. Let's assume that $P(a_1) = P(a_2)$ and make a pooled estimate, p. Then from formula (32)

$$v(p) = p(1-p)/b_1 + b_2 . \tag{40}$$

If we assume that $P(a_1) \neq P(a_2)$,

$$v(p_1) = p_1(1-p_1)/b_2 \tag{41}$$

$$\text{and} \quad v(p_2) = p_2(1-p_2)/b_1 \tag{42}$$

Then our estimate of the variance of $p_{1,2}$ from formula (33), assuming that we made a pooled estimate p from formula (33), is

$$v(p_{1,2}) = p^2 v(p) + p^2 v(p) \tag{43}$$

$$\text{or} \quad v(p_{1,2}) = 2\left[p^2 v(p)\right]$$

and if we did not make a pooled estimate, p, the variance formula is

$$v(p_{1,2}) = p_2^{\,2} v(p_1) + p_1^{\,2} v(p_2) \tag{44}$$

Inasmuch as $p_{1,2}$ is distributed binomially

$$v(q) = v(p_{1,2}) \tag{45}$$

and from formula (33) the variance of s can be estimated as follows

$$v(s) = M^2 v(q) \tag{46}$$

Then, our estimates of the variance of t and y would be as follows

$$v(t) = \frac{s^2}{(x/n)^4} v(x/n) + \frac{1}{(x/n)^2} v(s) \tag{47}$$

$$\text{and} \quad v(y) = \frac{(x/n)^2}{s^4} v(s) + \frac{1}{s^2} v(x/n) \tag{48}$$

By use of the basic formulas presented, $v(t)$ and $v(y)$ can be determined when more than two marks are applied.

Let's consider an example. We will use the same basic data, modified somewhat, which we borrowed from Ricker previously. Let's assume that 110 animals were marked originally with two marks and each mark had a probability of being lost of .10. Then $p_{1,2} = (.10)(.10) = .01$, $q = .99$, and $s = .99(110) = 109$. (If we had applied only one mark to the animal and p equaled .10, we would have had to mark 121 animals to expect s to equal 109.) Out of our sample of 57 animals we found $b_1 + b_2 = 103$ and $a_2 + a_1 = 10$. Then $p = 10/103$ or approximately .10 and

$$v(p) = .10(1-.10)/103$$

$$= .0087$$

and $$v(p_{1,2}) = 2\left[.10^2(.0087)\right]$$

$$= .000,174$$

and $$v(q) = .000,174.$$

Then,

$$v(s) = (110)^2\ .000174$$

$$= 2.1054$$

As previously, $v(x/n) = .001233$, then

$$v(t) = \frac{109^2}{.322^4}\ .001233 + \frac{1}{.322^2}\ 2.1054$$

$$= 1382.982$$

and $$v(y) = \frac{.322^2}{109^4}\ 2.1054 + \frac{1}{109^2}\ .001233$$

$$= .000,000,010,532,5$$

The approximate 95 percent confidence limits of t and y are

$$(\underline{t}, \overline{t}) = 339 \pm 2\sqrt{1382.982}$$

$$= 265 - 413$$

$$\text{and} \quad (\underline{y}, \bar{y}) = .00295 \pm 2\sqrt{.000,000,010,5}$$

$$= .002,745 - .003,155$$

Inverting the limits of y gives limits of $(\underline{t}, \bar{t})$ equal to 317-364. Actually the variance of s contributed very little to the variance ot t and y in the example presented here, which is shown by the fact that there was no measurable change in the confidence limits from that presented previously: however, this may not be necessarily true in all cases.

DISCUSSION

What has been presented in this report is only a preliminary consideration of the problem of mark loss. We would like to encourage others to consider this problem in more detail, especially the mathematical aspects of the problem and the determination of confidence limits.

The formulas for the determination of confidence limits and the variance of the estimates are only large sample approximations. However, the question arises, how large does the sample have to be for the approximation to hold and for the estimates to be approximately normally distributed? We don't have this answer, and surely this is worthy of further study. Also, which confidence limits, $(\underline{t}, \bar{t})$ or the inverted limits of $(\underline{y}, \bar{y})$ are best? In the example, the inverted limits of $(\underline{y}, \bar{y})$ gave the shortest limits, however an approximation cannot be chosen solely because it yields the shortest limits. This is also worthy of further study.

It should be noted that the formulas for the determination of variance are generalized. This was done on purpose so that they could be used with different sampling plans and assumptions. It should be emphasized that the formulas for the estimation of the variance of the estimate are only large sample approximations and in small samples they could under estimate the variance. Formula (34) for the variance of a quotient can be put into the form which Finney (1952) refers to as a "naive result". He suggests that for the use of formula (34) X must be at least 9 times its standard error for the setting of the 95 percent confidence limits. Finney presents formulas for the determination of fiducial limits of a quotient based on the t distribution and the Behrens - Fisher distribution which would be appropriate for small samples. For the assumptions, we have used in this report, fiducial limits based on the Behrens-Fisher distribution would be the most applicable.

There was one source of variation which we did not consider in the determination of the variance of the estimate. As pointed out previously, even if the proportion of animals losing mark 1 and mark 2 was known without error, $p_{1,2}$ is only the most likely proportion of animals losing their identity in the finite population of animals we have marked. However, if a relatively large number of animals are marked, i.e., if M is large, we believe that the contribution of this source of variation to the variance of the estimate will be negligible and can be ignored.

ACKNOWLEDGMENTS

We are indebted to Mr. Robert Murry and Mr. Lawrence Soileau, Louisiana Wild Life and Fisheries Commission for introducing us to the problem of mark loss. We have gained much from the suggestions and discussions with the above mentioned persons and with Mr. Scott Overton of the Cooperative Statistical Project, North Carolina State University; however, inasmuch as we have not always chosen to take their sound advice the responsibility for any errors rests solely with us.

LITERATURE CITED

Anderson, R. L. and T. A. Bancroft 1952. Statistical theory in research. McGraw-Hill Book Co., Inc., New York: xix + 399 pp.

Bailey, N. J. J. 1952. Improvements in the interpretation of recapture data. J. Animal Ecology, 21: 120-127.

Beverton, R. J. H. and S. J. Holt 1957. On the dynamics of exploited fish populations. Ministry of Agriculture, Fisheries and Food, Fishery Investigations, Series II, Vol. xix. Her Majesty's Stationery Office, London, England: 533 pp.

Dennett, D. and J. B. Kidd 1960. An analysis by tag returns of three years controlled hunting. Presented at the Fourteenth Annual Conf., S.E. Assoc. of Game and Fish Commissioners, 14 pp. (processed).

Feller, William 1957. An introduction to probability theory and its applications. Vol. 1, 2nd ed., John Wiley and Sons, Inc., New York: xv + 461 pp.

Finney, D. J. 1952. Statistical method in biological assay. Charles Griffin and Company, Limited, London, England: xix + 661 pp.

Goldberg, Samuel 1960. Probability an introduction. Prentice-Hall, Inc., Englewood Cliffs, N. J.: xiv + 322 pp.

Mainland, D., L. Herrera and M. I. Sutcliffe 1956. Statistical tables for use with binomial samples - contingency tests, confidence limits, and sample size estimates. Dept. of Medical Statistics, New York University College of Medicine, New York: xix + 83 pp.

Murry, R. E. 1960. Supplemental computations from recovery of marked deer on Red Dirt - 1959. In Annual Progress Report, Louisiana Wild Life and Fisheries Commission, Pittman-Robertson Project W-29-9 (Mimeo.).

Ricker, W. E. 1958. Handbook of computations for biological statistics of fish populations. Fish. Res. Board of Canada, Ottawa, Bull. 119:300 pp.

LITERATURE CITED

Anderson, R. L. and T. A. Bancroft 1952. Statistical theory in research. McGraw-Hill Book Co., Inc., New York: xix + 399 pp.

Bailey, N. J. J. 1952. Improvements in the interpretation of recapture data. J. Animal Ecology, 21: 120-127.

Beverton, R. J. H. and S. J. Holt 1957. On the dynamics of exploited fish populations. Ministry of Agriculture, Fisheries and Food, Fishery Investigations, Series II, Vol. xix. Her Majesty's Stationery Office, London, England: 533 pp.

Bennett, D. and J. B. Kidd 1960. An analysis by tax returns of three years controlled hunting. Presented at the Fourteenth Annual Conf., S.E. Assoc. of Game and Fish Commissioners, 14 pp. (processed).

Feller, William 1957. An introduction to probability theory and its applications. Vol. 1, 2nd ed., John Wiley and Sons, Inc., New York: xv + 461 pp.

Finney, D. J. 1952. Statistical method in biological assay. Charles Griffin and Company, Limited, London, England: xix + 661 pp.

Goldberg, Samuel 1960. Probability an introduction. Prentice-Hall, Inc., Englewood Cliffs, N. J.: xiv + 322 pp.

Mainland, D., L. Herrera and M. I. Sutcliffe 1956. Statistical tables for use with binomial samples - contingency tests, confidence limits, and sample size estimates. Dept. of Medical Statistics, New York University College of Medicine, New York: xix + 83 pp.

Murry, R. E. 1960. Supplemental computations from recovery of marked deer on Red Dirt - 1959. In Annual Progress Report, Louisiana Wild Life and Fisheries Commission, Pittman-Robertson Project W-29-9 (Mimeo.).

Ricker, W. E. 1958. Handbook of computations for biological statistics of fish populations. Fish. Res. Board of Canada, Ottawa, Bull. 119:300 pp.

APPENDIX I.

VARIANCE OF FUNCTION OF INDEPENDENT VARIATES

Let Z be a function of two variates, X and Y, i.e., $Z = F(X,Y)$. Then suppose two values of X and Y are taken which deviate from their means by the small amounts ΔX and ΔY, respectively. Then the deviation of ΔZ of the corresponding value of Z from its mean is

$$\Delta Z \simeq \frac{\partial Z}{\partial X} \Delta X + \frac{\partial Z}{\partial Y} \Delta Y \qquad \text{(a)}$$

From this we can determine the relations between the variance of Z, X and Y. Then V(Z), the variance of the resulting Z about its mean, will be $\sum (\Delta Z)^2/N$. If we square both sides of equation (a) we get

$$\Delta Z^2 = \left(\frac{\partial Z}{\partial X}\right)^2 \Delta X^2 + \left(\frac{\partial Z}{\partial Y}\right)^2 \Delta Y^2 + 2\left(\frac{\partial Z}{\partial X}\right)\left(\frac{\partial Z}{\partial Y}\right) (\Delta X)(\Delta Y)$$

Now, $\sum (\Delta X)^2/N$ and $\sum (\Delta Y)^2/N$ are the variances of X and Y, respectively, i.e., V(X) and V(Y), respectively. Then,

$$V(Z) = \left(\frac{\partial Z}{\partial X}\right)^2 \frac{\sum (\Delta X)^2}{N} + \left(\frac{\partial Z}{\partial Y}\right)^2 \frac{\sum (\Delta Y)^2}{N} + 2\left(\frac{\partial Z}{\partial X}\right)\left(\frac{\partial Z}{\partial Y}\right) \frac{\sum [(\Delta X)(\Delta Y)]}{N} \qquad \text{(b)}$$

However, since X and Y are independent their covariance will equal zero, then $\sum [(\Delta X)(\Delta Y)]/N$ will equal zero. Thus

$$2\left(\frac{\partial Z}{\partial X}\right)\left(\frac{\partial Z}{\partial Y}\right) \frac{\sum [(\Delta X)(\Delta Y)]}{N} \text{ will vanish}$$

and $$V(Z) = \left(\frac{\partial Z}{\partial X}\right)^2 V(X) + \left(\frac{\partial Z}{\partial Y}\right)^2 V(Z)$$

Now, if $Z = XY$, then $(\partial Z/\partial X) = Y$ and $(\partial Z/\partial Y) = X$ and

$$V(Z) = Y^2 V(X) + X^2 V(Y) \qquad \text{(c)}$$

Now, if $Z = Y/X$, then $(\partial Z/\partial X) = -Y/X^2$

and $(\partial Z/\partial Y) = 1$ and

$$V(Z) = \frac{Y^2}{X^4} V(X) + \frac{1}{X^2} V(Y) \qquad \text{(d)}$$

Formula (c) determines the variance of a product while formula (d) determines the variance of a quotient. These are large sample approximations.

3 0112 106073361

APPENDIX I.

VARIANCE OF FUNCTION OF INDEPENDENT VARIATES

Let Z be a function of two variates, X and Y, i.e., $Z = F(X,Y)$. Then suppose the values of X and Y are taken which deviate from their means by the small amounts ΔX and ΔY, respectively. Then the deviation of ΔZ of the corresponding value of Z from its mean is

$$\Delta Z = \frac{\partial Z}{\partial X}\Delta X + \frac{\partial Z}{\partial Y}\Delta Y \qquad (a)$$

From this we can determine the relations between the variance of Z, X and Y. Then $V(Z)$, the variance of the resulting Z about its mean, will be $\sum (\Delta Z)^2/n$. If we square both sides of equation (a) we get

$$\Delta Z^2 = \left(\frac{\partial Z}{\partial X}\right)^2 \Delta X^2 + \left(\frac{\partial Z}{\partial Y}\right)^2 \Delta Y^2 + 2\left(\frac{\partial Z}{\partial X}\right)\left(\frac{\partial Z}{\partial Y}\right)(\Delta X)(\Delta Y)$$

Now, $\sum (\Delta X)^2/n$ and $\sum (\Delta Y)^2/n$ are the variances of X and Y, respectively, i.e., $V(X)$ and $V(Y)$, respectively. Then,

$$V(Z) = \left(\frac{\partial Z}{\partial X}\right)^2 \frac{\sum (\Delta X)^2}{n} + \left(\frac{\partial Z}{\partial Y}\right)^2 \frac{\sum (\Delta Y)^2}{n} + 2\left(\frac{\partial Z}{\partial X}\right)\left(\frac{\partial Z}{\partial Y}\right)\frac{\sum [(\Delta X)(\Delta Y)]}{n} \qquad (b)$$

However, since X and Y are independent their variations will equal zero, that $\sum [(\Delta X)(\Delta Y)]/n$ will equal zero. Thus

$$= \left(\frac{\partial Z}{\partial X}\right)\left(\frac{\partial Z}{\partial Y}\right)\frac{\sum [(\Delta X)(\Delta Y)]}{n} \quad \text{will equal}$$

and $\quad V(Z) = \left(\frac{\partial Z}{\partial X}\right)^2 V(X) + \left(\frac{\partial Z}{\partial Y}\right)^2 V(Y)$

Now, if $Z = XY$, then $(\partial Z/\partial X) = Y$ and $(\partial Z/\partial Y) = X$ and

$$V(Z) = Y^2 V(X) + X^2 V(Y) \qquad (c)$$

Now, if $Z = Y/X$, then $(\partial Z/\partial X) = -Y/X^2$

and $\quad (\partial Z/\partial Y) = 1$ and

$$V(Z) = \frac{Y^2}{X^4} V(X) + \frac{1}{X^2} V(Y) \qquad (d)$$

Formula (c) determines the variance of a product while formula (d) determines the variance of a quotient. These are large sample approximations.